A Note on the Itô and Stratonovich Stochastic Relative Degree and Normal Form

Alberto Mellone and Giordano Scarciotti

***Abstract*— In this note we compare two notions of stochastic relative degree. Specifically, we consider nonlinear stochastic systems defined by the same stochastic differential equation and interpreted in either Itô's or Stratonovich's sense. We then recall the Itô stochastic relative degree and we introduce the concept of Stratonovich stochastic relative degree and Stratonovich normal form. We show, by means of examples, that the stochastic relative degrees arising from the two different interpretations of the same stochastic differential equations are, in general, different. We finally point out that this discrepancy can be eliminated through conversion formulas between Itô and Stratonovich integrals.**

I. INTRODUCTION

The analysis and control of stochastic systems are relevant topics in control engineering. The importance of studying stochastic systems is twofold. Firstly, this framework allows for modelling unknown dynamics through stochastic differential equations when model approximations yield stochastic parameter uncertainties [1]. Secondly, stochastic systems have several applications ranging from classical control topics like the optimal stopping problem to production planning, finance and technology diffusion [1], [2].

A fundamental fact in the theory of stochastic differential equations is that there is no unique interpretation of the stochastic integral [1], [3]. This discrepancy yields different solutions of the same stochastic differential equation, when this is interpreted in different ways. The two most common interpretations of the stochastic integral are Itô's and Stratonovich's. Itô's interpretation is causal, representing that at any given time the value of the stochastic integral is dependent upon present or past history of the integrand. This interpretation is the prevalent approach in mathematics and the solution of stochastic differential equations interpreted in Itô's sense obeys specific calculus rules that are different from ordinary, deterministic calculus. On the other hand, Stratonovich's interpretation of the stochastic integral is prevalent in physics and, although it satisfies the deterministic calculus rules, it requires the knowledge of the integrand at a future time, thus it is non-causal. Having said that multiple interpretations of the "formally identical" stochastic integral yield different solutions, conversion formulas can be used to relate these different solutions [3], [4]. Therefore a problem in the stochastic framework can be solved employing the most convenient interpretation (for example the causality of the Itô integral or the deterministic calculus rules of the Stratonovich integral) and retrieving the solution in a different interpretation through conversion formulas.

This study addresses the definition of the stochastic relative degree of nonlinear stochastic systems when the underlying stochastic differential equation is interpreted either in Itô's or Stratonovich's sense. We consider the general case of stochastic differential equations where both drift and diffusion terms are present and are functions of the state and of the control input. We show that for the formally identical system, the two interpretations of the stochastic differential equations yield in general different relative degrees. For the sake of providing a full motivating picture, we recall that the definition of the relative degree is instrumental for the introduction of a "normal form" of the nonlinear dynamics, *i.e.* a change of coordinates which expresses the dynamics of the system in a simpler way. In control theory the definition of a normal form is employed to control the system, *e.g.* via feedback linearisation or high-gain feedback [5]. For the sake of completeness we mention that the term "normal form" is a generic term which is used in theoretical fields to indicate a special way in which the system is written. Another "normal form", not related to control, is given for example in [6] and [7], where it is defined for purely diffusive processes interpreted in Stratonovich's sense. Other examples of coordinate changes for stochastic systems can be found, *e.g.*, in [8] and [9].

A normal form for nonlinear stochastic systems interpreted in Itô's sense for control purposes was introduced in [10], where the authors also solve the linearisation via feedback problem. In [11] the normal form is employed to solve the problem of local stabilisation and tracking in the unrealistic assumption that the noise can be measured. By removing this hypothesis and approximating the noise, the same problems are solved in an approximate way in [12].

In this note we define the Stratonovich stochastic relative degree and normal form and then we compare the two definitions of Itô and Stratonovich stochastic relative degree. We then present two examples which illustrate the fact that the two interpretations of the stochastic differential equations yield different stochastic relative degrees.

The rest of the paper is organised as follows. In Section II we recall some preliminary notions on stochastic systems, differential operators and stochastic calculus. In Section III we recall the notion of Itô stochastic relative degree introduced in [12]. We then introduce Stratonovich differential operators, the Stratonovich stochastic relative degree and the Stratonovich normal form. In Section IV we present two examples to illustrate the theory. Finally, Section V contains

Alberto Mellone and Giordano Scarciotti are with the Department of Electrical and Electronic Engineering, Imperial College London, London, SW7 2AZ, UK. {a.mellone18,g.scarciotti}@imperial.ac.uk

some concluding remarks.

Notation. The symbol $\mathbb{Z}$ denotes the set of integer numbers, while $\mathbb{R}$ and $\mathbb{C}$ denote the fields of real and complex numbers, respectively; by adding the subscript "< 0" ("≥ 0", "0") to any symbol indicating a set of numbers, we denote that subset of numbers with negative (non-negative, zero) real part. The symbol ∂_x^n is used as a shorthand for the operator $\partial^n/\partial x^n$, while $\alpha^{(n)}$ indicates the n-th time derivative of α. $(\nabla, \mathfrak{A}, \mathfrak{P})$ is a probability space given by the set ∇, the σ-algebra $\mathfrak{A}$ defined on ∇ and the probability measure $\mathfrak{P}$ on the measurable space $(\nabla, \mathfrak{A})$. A *stochastic process* with state space $\mathbb{R}^n$ is a family $\{x_t,\, t \in \mathbb{R}\}$ of $\mathbb{R}^n$-valued random variables, *i.e.* for every fixed $t \in \mathbb{R}$, $x_t(\cdot)$ is an $\mathbb{R}^n$-valued random variable and, for every fixed $w \in \nabla$, $x.(w)$ is an $\mathbb{R}^n$-valued function of time [3, Section 1.8]. For ease of notation, we often indicate a stochastic process $\{x_t,\, t \in \mathbb{R}\}$ simply with x_t (this is common in the literature, see *e.g.* [3]). With a slight abuse of notation, any subscript different from the symbol "t" indicates the corresponding component of the vector x_t, *e.g.* x_i is the i-th component of the vector x_t. All mappings appearing as integrands in stochastic integrals are assumed to be integrable in the corresponding sense, namely Itô's (see, *e.g.*, [1, Definition 3.1.4]) or Stratonovich's (equivalent to deterministic integrability).

II. Preliminaries

In this section we first recall the theory of generalised stochastic processes and define differential operators that will be used in the remainder of the paper. We then briefly illustrate the main differences between the Itô and Stratonovich stochastic integrals.

A. Stochastic processes and differential operators

Let $C_0^\infty(\mathbb{R})$ be the space of all infinitely differentiable functions on $\mathbb{R}$ with compact support [13, Definition 1.2.1]. The following definition characterises the notion of *distribution*, or, equivalently, of *generalised function*.

Definition 1. [14, Definition 3.1] Let X be an open subset of $\mathbb{R}$. A *distribution on* X is a linear form ψ on $C_0^\infty(\mathbb{R})$ that is also continuous in the sense that

$$\lim_{j\to\infty} \psi(\varphi_j) = \psi(\varphi) \qquad \text{as} \qquad \lim_{j\to\infty} \varphi_j = \varphi \quad \text{in} \quad C_0^\infty(\mathbb{R}).$$

If f is a continuous and differentiable function, then by partial integration we get

$$\int_{\mathbb{R}} \dot{f}\varphi \, dt = -\int_{\mathbb{R}} f\dot{\varphi} \, dt, \quad \forall \varphi \in C_0^\infty(\mathbb{R}).$$

By analogy, for any distribution ψ, its distributional derivative $\dot{\psi}$ is defined as the distribution that satisfies

$$\dot{\psi}(\varphi) = -\psi(\dot{\varphi}), \quad \forall \varphi \in C_0^\infty(\mathbb{R}),$$

see [13, Definition 3.1.1]. Note also that generalised functions have derivatives of all order, which are generalised functions as well.

Definition 2. [3, Section 3.2] A generalised stochastic process is a random generalised function in the sense that a random variable $\psi(\varphi)$ is assigned to every $\varphi \in C_0^\infty$, where ψ is, with probability 1, a generalised function.

The distributional derivative of the Brownian motion is always defined [3, Section 3.2]. In particular, the generalised stochastic process given by such a derivative has zero mean value and covariance function given by the generalised function $\delta(t-s)$, $t, s \in \mathbb{R}$, *i.e.* the Dirac delta. Consequently, the derivative of the Brownian motion is the generalised *white noise* [3, Section 3.2]. In the remainder, with a slight abuse of notation, we refer to the generalised white noise omitting the attribute "generalised" and we denote it by simply ξ_t, with $\xi_t = \dot{\mathcal{W}}_t$, where $\mathcal{W}_t$ is a standard Wiener process, also known as Brownian motion. It should be emphasised that the just mentioned time derivative is meant in the sense of distributions, and not as the limit of the difference quotient as the increment tends to zero, which instead applies to differentiable functions in the classical sense.

We now introduce two differential operators that will be used in the remainder of the paper.

Recall that the derivative of h along the vector field f, which is called Lie derivative and is indicated with the symbol $\mathcal{L}_f\, h$, is defined as [5]

$$\mathcal{L}_f\, h(x) = \partial_x[h]\; f(x) = \sum_{i=1}^{n} \frac{\partial h}{\partial x_i} f_i(x).$$

We indicate the derivative of h first along the vector field f and then along the vector field g as $\mathcal{L}_g\, \mathcal{L}_f\, h(x) = \partial_x[\mathcal{L}_f\, h]\;\; g(x)$. We use the recursive relation $\mathcal{L}_f^k\, h(x) = \partial_x[\mathcal{L}_f^{k-1}\, h]\; f(x)$, with $\mathcal{L}_f^0\, h(x) = h(x)$, to indicate the k-th differentiation of h along f.

The second operator, which indicates the second derivative of h along the vector fields f and g, is defined as [10]

$${}^g\mathcal{G}_f h(x) = g(x)^\top \partial_x^2[h]\; f(x) = \sum_{j=1}^{n} g_j(x) \sum_{i=1}^{n} \frac{\partial^2 h}{\partial x_j \partial x_i} f_i(x).$$

Similarly to the Lie derivative, we use the notation ${}^b\mathcal{G}_a\, {}^g\mathcal{G}_f h(x) = b(x)^\top \partial_x^2[\,{}^g\mathcal{G}_f h]\;\; a(x)$, and ${}^g\mathcal{G}_f^k h(x) = g(x)^\top \partial_x^2[\,{}^g\mathcal{G}_f^{k-1} h]\; f(x)$, to indicate the reiterated operations.

B. The Itô and Stratonovich stochastic integrals

For the sake of readability, we briefly recall the main differences between the two main interpretations of the stochastic integral.

Given a map $l : \mathbb{R}_{\geq 0} \times \nabla \to \mathbb{R}^n$, a stochastic integral is a stochastic process of the form

$$X(t_n, w) = X_{t_0} + \int_{t_0}^{t_n} l(\tau, w) d\mathcal{W}_\tau \tag{1}$$

which can be approximated as [15, Section 2.1]

$$X(t_n, w) \approx X_{t_0} + \sum_{k=1}^{n} l(\tau_{k-1}, w)(\mathcal{W}_{t_k} - \mathcal{W}_{t_{k-1}}).$$

The value τ_{k-1} can be arbitrarily chosen in the interval $[t_{k-1}, t_k]$, thus yielding different intepretations and properties: if $\tau_{k-1} = t_{k-1}$, then the Itô integral is obtained, which

has the advantage of being causal but it does not obey the standard rules of calculus; if $\tau_{k-1} = (t_k + t_{k-1})/2$, then the Stratonovich integral is obtained [16], which has the advantage of obeying the rules of deterministic integrals but it is non-causal. The shorthand notation

$$dX_t = l(t, w)d\mathcal{W}_t$$

indicates the Itô stochastic differential equation satisfied by X_t and is used to refer to (1) in Itô's sense, whereas

$$dX_t = l(t, w) \circ d\mathcal{W}_t$$

indicates the Stratonovich stochastic differential equation satisfied by X_t and is used to refer to (1) in Stratonovich's sense.

Let $f : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ and $l : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$. Consider the stochastic process $X_t \in \mathbb{R}^n$ defined by

$$dX_t = f(t, X_t)dt + l(t, X_t)d\mathcal{W}_t, \tag{2}$$

and let $Y_t = g(t, X_t)$, with $g : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ be a mapping which is differentiable in t and twice-differentiable in X_t. Then, by Itô's formula [3], the stochastic differential equation satisfied by Y_t is

$$dY_t = \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial X_t}dX_t + \frac{1}{2}dX_t^\top \frac{\partial^2 g}{\partial X_t^2}dX_t,$$

where in the second order term the stochastic calculus rules $dt^2 = dtd\mathcal{W}_t = 0$ and $d\mathcal{W}_t^2 = dt$ must be applied. Note that the previous expression can be interpreted as the chain rule for processes interpreted in Itô's sense.

On the other hand, when X_t satisfies

$$dX_t = f(t, X_t)dt + l(t, X_t) \circ d\mathcal{W}_t, \tag{3}$$

and $Y_t = g(t, X_t)$, the Stratonovich's chain rule is the usual deterministic chain rule, and the stochastic differential equation satisfied by Y_t is

$$dY_t = \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial X_t}dX_t.$$

Note that even though (2) and (3) look formally the same, the two interpretations correspond to two different modelling assumptions of the underlying noise affecting the system. In Itô's sense the noise is generalised white noise. In Stratonovich's sense the noise is a sufficiently smooth approximation of the white noise. It is this fundamental difference in the regularity of the noise that gives rise to the two different mathematical frameworks.

We are now ready to provide the definitions of stochastic relative degree when the nonlinear stochastic differential equation defining the system is intended in either Itô's or Stratonovich's sense.

III. Stochastic Relative Degree in Itô's and Stratonovich's Sense

In this section we first recall the definition of Itô stochastic relative degree from [10]. Then we provide the definition of Stratonovich stochastic relative degree and Stratonovich normal form and compare the two mathematical frameworks.

A. The Itô stochastic relative degree [10]

We first consider the nonlinear single-input, single-output stochastic system expressed in the shorthand integral notation by

$$\begin{aligned} dx_t &= (f(x_t) + g(x_t)u)dt + (l(x_t) + m(x_t)u)d\mathcal{W}_t, \\ y_t &= h(x_t), \end{aligned} \tag{4}$$

with $x_t \in \mathbb{R}^n$, $u \in \mathbb{R}$, $y_t \in \mathbb{R}$ and $f : \mathbb{R}^n \to \mathbb{R}^n$, $g : \mathbb{R}^n \to \mathbb{R}^n$, $l : \mathbb{R}^n \to \mathbb{R}^n$, $m : \mathbb{R}^n \to \mathbb{R}^n$, $h : \mathbb{R}^n \to \mathbb{R}$ *smooth* functions, *i.e.* they admit continuous partial derivatives of any order. We assume that, for a fixed initial condition $x_{t=0}$, the solution of (4) is unique. Note that, in the light of the considerations of Section II, system (4) can be rewritten in the following differential notation

$$\dot{x}_t = f(x_t) + g(x_t)u + (l(x_t) + m(x_t)u)\xi_t, \quad y_t = h(x_t). \tag{5}$$

As recalled in Section II, when ξ_t is (generalised) white noise, as in this case, then the differential equation (5) is equivalent to the integral equation (4) when the latter is interpreted in Itô's sense [3, Section 10.3]. Given the equivalence of the two representations in the framework of generalised stochastic processes, in the remainder of the paper equations (4) and (5) are used interchangeably, as convenient, to refer to the same nonlinear stochastic system intended in Itô's sense.

Let ${}^l\mathcal{S}_f h$ be the Itô stochastic Lie derivative of h along the drift vector field f and diffusion vector field l. This operator is defined as [10]

$${}^l\mathcal{S}_f h(\xi_t, x) = \mathcal{L}_f\, h(x) + \mathcal{L}_l\, h(x)\xi_t + \frac{1}{2}\,{}^l\mathcal{G}_l h(x).$$

Similarly to the deterministic Lie derivative, if ${}^l\mathcal{S}_f h(\xi_t, x) = {}^l\mathcal{S}_f h(x)$ is a deterministic expression, *i.e.* the white noise does not appear explicitly, we use the notation ${}^l\mathcal{S}_f^2 h(\xi_t, x) = {}^l\mathcal{S}_f\,{}^l\mathcal{S}_f h(\xi_t, x)$. Iteratively, if ${}^l\mathcal{S}_f^{k-1} h(\xi_t, x) = {}^l\mathcal{S}_f^{k-1} h(x)$ is deterministic, then ${}^l\mathcal{S}_f^k h(\xi_t, x) = {}^l\mathcal{S}_f\,{}^l\mathcal{S}_f^{k-1} h(\xi_t, x)$, with ${}^l\mathcal{S}_f^0 h(x) = h(x)$ by definition. Finally, we define the operator

$${}^m_g\mathcal{A}_l h(\xi_t, x) = \mathcal{L}_g\, h(x) + \mathcal{L}_m\, h(x)\xi_t + {}^m\mathcal{G}_l h(x).$$

By using Itô's formula, it is easy to see that the first derivative of the output of system (5) is given by

$$y_t^{(1)} = {}^l\mathcal{S}_f h(\xi_t, x_t) + {}^m_g\mathcal{A}_l h(\xi_t, x_t)u + \frac{1}{2}\,{}^m\mathcal{G}_m h(x_t)u^2.$$

We now recall the definition of stochastic relative degree when the stochastic differential equation (4) is interpreted in Itô's sense.

Definition 3. [10] (Itô Stochastic Relative Degree) Assume that there exists $\bar{r}$ such that

$$\mathcal{L}_l\,{}^l\mathcal{S}_f^k h(\xi_t, x) = 0, \qquad \forall\, k \in \{0, ..., \bar{r} - 2\},$$

and for all x in a neighbourhood of $\bar{x}$. System (5) is said to have *Itô stochastic relative degree* r at a point $\bar{x}$ if $\bar{r} = r$ and

1) $\mathcal{L}_g\,{}^l\mathcal{S}_f^k h(\xi_t, x) + {}^m\mathcal{G}_l\,{}^l\mathcal{S}_f^k h(\xi_t, x) = 0$ and $\mathcal{L}_m\,{}^l\mathcal{S}_f^k h(\xi_t, x) = 0$ and ${}^m\mathcal{G}_m\,{}^l\mathcal{S}_f^k h(\xi_t, x) = 0$ for all x in a neighborhood of $\bar{x}$ and all $k \in \{0, ..., r - 2\}$.

2) $\mathcal{L}_g\ {}^{l}\mathcal{S}_f^{r-1}h(\xi_t,\bar{x}) + {}^{m}\mathcal{G}_l\ {}^{l}\mathcal{S}_f^{r-1}h(\xi_t,\bar{x}) \neq 0$ or $\mathcal{L}_m\ {}^{l}\mathcal{S}_f^{r-1}h(\xi_t,\bar{x}) \neq 0$ or ${}^{m}\mathcal{G}_m\ {}^{l}\mathcal{S}_f^{r-1}h(\xi_t,\bar{x}) \neq 0$.

It is easy to observe that the definition of stochastic relative degree extends the definition of the deterministic counterpart, in that it is equal to the order of the derivative of the output in which the input explicitly appears.

B. The Stratonovich stochastic relative degree

We now consider the nonlinear single-input, single-output stochastic system expressed by

$$\begin{aligned} dx_t &= (f(x_t)+g(x_t)u)dt + (l(x_t)+m(x_t)u)\circ d\mathcal{W}_t,\\ y_t &= h(x_t), \end{aligned} \tag{6}$$

and we again assume that, for a fixed initial condition $x_{t=0}$, the solution of (6) is unique. Note that the symbol $\circ$ denotes that the stochastic differential equation is meant in Stratonovich's sense. As recalled in Section II, it is known that under appropriate assumptions (see [3, Section 10.3] and references therein) the stochastic differential equation (6) can be rewritten in the form

$$\dot{x}_t = f(x_t)+g(x_t)u+(l(x_t)+m(x_t)u)\zeta_t, \tag{7}$$

where $\zeta_t \in \mathbb{R}$ is a sufficiently smooth stochastic process which is an approximation of the white noise. To streamline the presentation we also introduce the vector $\boldsymbol{\zeta}_t = \begin{bmatrix}\zeta_t & \zeta_t^{(1)} & \dots & \zeta_t^{(n-1)}\end{bmatrix}^\top$, where the element $\zeta_t^{(i)}$ inidcates the i-th time derivative of the process ζ_t.

We now introduce the Stratonovich stochastic Lie derivative of h along the drift vector field f and the diffusion vector field l as

$${}^{l}\widehat{\mathcal{S}}_f h(\boldsymbol{\zeta}_t,x) = \mathcal{L}_f\, h(x) + \mathcal{L}_l\, h(x)\zeta_t,$$

and, in analogy with previous definitions, ${}^{l}\widehat{\mathcal{S}}_f^{\,k} h(\boldsymbol{\zeta}_t,x) = {}^{l}\widehat{\mathcal{S}}_f\ {}^{l}\widehat{\mathcal{S}}_f^{\,k-1} h(\boldsymbol{\zeta}_t,x)$, with ${}^{l}\widehat{\mathcal{S}}_f^{\,0} h(\boldsymbol{\zeta}_t,x) = h(x)$. Applying the Stratonovich rules for the differentiation of a stochastic process, it is easy to see that

$$y_t^{(1)} = {}^{l}\widehat{\mathcal{S}}_f h(\boldsymbol{\zeta}_t,x_t) + {}^{m}\widehat{\mathcal{S}}_g h(\boldsymbol{\zeta}_t,x_t)u. \tag{8}$$

We now provide the definition of stochastic relative degree in Stratonovich's sense.

Definition 4. (Stratonovich Stochastic Relative Degree) System (7) is said to have *Stratonovich stochastic relative degree* r at a point $\bar{x}$ if

1) ${}^{m}\widehat{\mathcal{S}}_g\ {}^{l}\widehat{\mathcal{S}}_f^{\,k} h(\boldsymbol{\zeta}_t,x) = 0$ for all x in a neighborhood of $\bar{x}$ and all $k \in \{0,\dots,r-2\}$.
2) ${}^{m}\widehat{\mathcal{S}}_g\ {}^{l}\widehat{\mathcal{S}}_f^{\,r-1} h(\boldsymbol{\zeta}_t,\bar{x}) \neq 0$.

Again, the Stratonovich stochastic relative degree is the number of time derivatives of the output, taken applying Stratonovich's differentiation rule, such that the control input appears. To see this, assume that the stochastic relative degree of system (7) is higher than one at a point $\bar{x}$. Then ${}^{m}\widehat{\mathcal{S}}_g h(\boldsymbol{\zeta}_t,x_t) = 0$ in (8) for all x in a neighbourhood of $\bar{x}$ and

$$y_t^{(1)} = {}^{l}\widehat{\mathcal{S}}_f h(\boldsymbol{\zeta}_t,x_t).$$

We now compute the second derivative of the output, yielding

$$y_t^{(2)} = {}^{l}\widehat{\mathcal{S}}_f^{\,2} h(\boldsymbol{\zeta}_t,x_t) + {}^{m}\widehat{\mathcal{S}}_g\ {}^{l}\widehat{\mathcal{S}}_f h(\boldsymbol{\zeta}_t,x_t)u + \frac{\partial y_t^{(1)}}{\partial \boldsymbol{\zeta}_t}\dot{\boldsymbol{\zeta}}_t,$$

which, under the assumption that the stochastic relative degree at $\bar{x}$ is higher than two, reduces to

$$y_t^{(2)} = {}^{l}\widehat{\mathcal{S}}_f^{\,2} h(\boldsymbol{\zeta}_t,x_t) + \frac{\partial y_t^{(1)}}{\partial \boldsymbol{\zeta}_t}\dot{\boldsymbol{\zeta}}_t$$

in a neighbourhood of $\bar{x}$. To streamline the presentation, let $\lambda_1(\boldsymbol{\zeta}_t,x_t) = \frac{\partial y_t^{(1)}}{\partial \boldsymbol{\zeta}_t}\dot{\boldsymbol{\zeta}}_t$, and iteratively, $\lambda_k(\boldsymbol{\zeta}_t,x_t) = \frac{\partial y_t^{(k)}}{\partial \boldsymbol{\zeta}_t}\dot{\boldsymbol{\zeta}}_t + {}^{l}\widehat{\mathcal{S}}_f\lambda_{k-1}(\boldsymbol{\zeta}_t,x_t)$ for $1 \leq k \leq r-1$ and $\lambda_0(\boldsymbol{\zeta}_t,x_t) = 0$. Repeating the differentiation procedure, under the assumption that the Stratonovich stochastic relative degree of system (7) is $r \leq n$, yields

$$y_t^{(k)} = {}^{l}\widehat{\mathcal{S}}_f^{\,k} h(\boldsymbol{\zeta}_t,x_t) + \lambda_{k-1}(\boldsymbol{\zeta}_t,x_t)$$

near $\bar{x}$ for all $k \in \{1,\dots,r-1\}$. Moreover, if, at time $\bar{t}$, $x_{\bar{t}} = \bar{x}$ holds, then

$$y_{\bar{t}}^{(r)} = {}^{l}\widehat{\mathcal{S}}_f^{\,r} h(\boldsymbol{\zeta}_{\bar{t}},\bar{x}) + {}^{m}\widehat{\mathcal{S}}_g\ {}^{l}\widehat{\mathcal{S}}_f^{\,r-1} h(\boldsymbol{\zeta}_{\bar{t}},\bar{x})u + \lambda_{r-1}(\boldsymbol{\zeta}_t,x_t)$$

with the input u explicitly appearing in $y_{\bar{t}}^{(r)}$.

Having provided the definition of Stratonovich relative degree, it is now natural to define a normal form for nonlinear stochastic systems interpreted in Stratonovich's sense.

Proposition 1. Suppose that system (7) has stochastic relative degree r at $\bar{x}$. Set

$$\begin{aligned} \widehat{\phi}_1(\boldsymbol{\zeta}_t,x) &= h(x),\\ \widehat{\phi}_2(\boldsymbol{\zeta}_t,x) &= {}^{l}\widehat{\mathcal{S}}_f h(x) + \lambda_0(\boldsymbol{\zeta}_t,x),\\ &\vdots\\ \widehat{\phi}_r(\boldsymbol{\zeta}_t,x) &= {}^{l}\widehat{\mathcal{S}}_f^{\,r-1} h(x) + \lambda_{r-2}(\boldsymbol{\zeta}_t,x). \end{aligned}$$

If $r < n$, then there exist smooth functions $\widehat{\phi}_{r+1}(\boldsymbol{\zeta}_t,x),\dots,\widehat{\phi}_n(\boldsymbol{\zeta}_t,x)$, with $\widehat{\phi}_j \in \mathbb{R}$ for all $j \in \{r+1,..,n\}$, such that the Jacobian of the mapping

$$\widehat{\Phi}(\boldsymbol{\zeta}_t,x) = \begin{bmatrix}\widehat{\phi}_1(\boldsymbol{\zeta}_t,x) & \widehat{\phi}_2(\boldsymbol{\zeta}_t,x) & \dots & \widehat{\phi}_n(\boldsymbol{\zeta}_t,x)\end{bmatrix}^\top$$

is invertible at $\bar{x}$ almost surely, thus defining a stochastic coordinate transformation in a neighbourhood of $\bar{x}$. Then the state-space representation of system (7) in the transformed state $z_t = \widehat{\Phi}(\boldsymbol{\zeta}_t,x_t)$ is

$$\begin{aligned} \dot{z}_i &= z_{i+1}, && i = 1,\dots,r-1,\\ \dot{z}_r &= \widehat{c}(\boldsymbol{\zeta}_t,z_t) + \widehat{b}(\boldsymbol{\zeta}_t,z_t)u,\\ \dot{z}_j &= \widehat{p}_j(\boldsymbol{\zeta}_t,z_t) + \widehat{q}_j(\boldsymbol{\zeta}_t,z_t)u, && j = r+1,\dots,n,\\ y_t &= z_1, \end{aligned}$$

where

$$\begin{aligned}
\widehat{c}(\boldsymbol{\zeta}_t, z_t) &= {}^{l}\widehat{\mathcal{S}}_f^{\,r} h(\boldsymbol{\zeta}_t, \widehat{\Phi}^{-1}(\boldsymbol{\zeta}_t, z_t)) + \lambda_{r-1}(\boldsymbol{\zeta}_t, \widehat{\Phi}^{-1}(\boldsymbol{\zeta}_t, z_t)),\\
\widehat{b}(\boldsymbol{\zeta}_t, z_t) &= {}^{m}\widehat{\mathcal{S}}_g\, {}^{l}\mathcal{S}_f^{r-1} h(\boldsymbol{\zeta}_t, \widehat{\Phi}^{-1}(\boldsymbol{\zeta}_t, z_t)),\\
\widehat{p}_j(\boldsymbol{\zeta}_t, z_t) &= {}^{l}\widehat{\mathcal{S}}_f \widehat{\phi}_j(\boldsymbol{\zeta}_t, \widehat{\Phi}^{-1}(\boldsymbol{\zeta}_t, z_t)) + \frac{\partial \widehat{\phi}_j}{\partial \boldsymbol{\zeta}_t}\dot{\boldsymbol{\zeta}}_t,\\
\widehat{q}_j(\boldsymbol{\zeta}_t, z_t) &= {}^{m}\widehat{\mathcal{S}}_g \widehat{\phi}_j(\boldsymbol{\zeta}_t, \widehat{\Phi}^{-1}(\boldsymbol{\zeta}_t, z_t)).
\end{aligned}$$

Note that, similarly to the deterministic case (see [5]), it might be possible to find smooth functions $\widehat{\phi}_{r+1}, ..., \widehat{\phi}_n$ such that the dynamics of the last $n-r$ transformed coordinates is independent of the input u, *i.e.* $\widehat{q}_j(\cdot, z_t) \equiv 0$, for all $j \in \{r+1, ..., n\}$, in a neighbourhood of $\widehat{\Phi}(\boldsymbol{\zeta}_t, \bar{x})$. This observation motivates the next definition.

Definition 5. (Stratonovich Stochastic Normal Form) Let x_t be the unique solution of (7) and $z_t = \widehat{\Phi}(\boldsymbol{\zeta}_t, x_t)$ be almost surely a local diffeomorphism in a subset U of $\mathbb{R}^n$ such that

$$\begin{aligned}
\dot{z}_i &= z_{i+1}, && i = 1, ..., r-1,\\
\dot{z}_r &= \widehat{c}(\boldsymbol{\zeta}_t, z_t) + \widehat{b}(\boldsymbol{\zeta}_t, z_t)u,\\
\dot{z}_j &= \widehat{p}_j(\boldsymbol{\zeta}_t, z_t), && j = r+1, ..., n,\\
y_t &= z_1.
\end{aligned} \tag{9}$$

System (9) is said to be the *Stratonovich stochastic normal form* of system (7).

For comparison, recall [10, Definition 4] that in Itô's sense one can define a local change of coordinates $z_t = \Phi(x_t)$ such that

$$\begin{aligned}
\dot{z}_i &= z_{i+1}, && i = 1, ..., r-1,\\
\dot{z}_r &= c(\xi_t, z_t) + b(\xi_t, z_t)u + a(z_t)u^2,\\
\dot{z}_j &= p_j(\xi_t, z_t), && j = r+1, ..., n,\\
y_t &= z_1,
\end{aligned}$$

with the expression of a, b, c, p_j omitted (these are available in [10]).

IV. Examples

In this section we present some examples that illustrate how the two different interpretations of the stochastic integral impact the relative degree of nonlinear stochastic systems.

Example 1. (Larger Stratonovich stochastic relative degree) In this first example we consider two systems having the same drift and diffusion vector fields and we show that when the stochastic differential equation is intended in Itô's sense, the stochastic relative degree is lower than the Stratonovich counterpart.

Let $x = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^\top$ and

$$f(x) = \begin{bmatrix} e^{-x_1}(x_3 - x_2) \\ x_2 \\ x_1 \end{bmatrix}, \quad g(x) = \begin{bmatrix} -e^{-x_1}\cos x_2 \\ \cos x_2 \\ 0 \end{bmatrix}, \tag{10}$$

$$l(x) = \begin{bmatrix} e^{-x_1}(1+x_1) \\ -(1+x_1) \\ 0 \end{bmatrix}, \; m(x) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \; h(x) = e^{x_1} + x_2. \tag{11}$$

We first consider the stochastic system described by (4), *i.e.* when the stochastic differential equation is intended in Itô's sense. Computing the differentials of the output y_t using Itô's formula yields

$$dy_t = \begin{bmatrix} e^{x_1} & 1 & 0 \end{bmatrix} dx_t + \frac{1}{2} dx_t^\top \begin{bmatrix} e^{x_1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} dx_t,$$

which simplifies to

$$y_t^{(1)} = x_3 + \frac{1}{2} e^{-x_1}(1 + x_1)^2.$$

As the control input does not appear in the previous expression, the Itô stochastic relative degree, if defined, is higher than one. Differentiating the output a second time yields

$$dy_t^{(1)} = \gamma_1(x_t)dt + \gamma_2(x_t)d\mathcal{W}_t + \gamma_3(x_t)u dt,$$

with

$$\begin{aligned}
\gamma_1(x_t) &= \frac{1}{2} e^{-2x_1}(1 - x_1^2)(x_3 - x_2)\\
&\quad + \frac{1}{4} e^{-3x_1}(1 + x_1)^2(x_1^2 - 2x_1 - 1) + x_1,\\
\gamma_2(x_t) &= \frac{1}{2} e^{-2x_1}(1 - x_1^2)(1 + x_1),\\
\gamma_3(x_t) &= \frac{1}{2} e^{-2x_1}(1 - x_1^2)\cos x_2.
\end{aligned}$$

Therefore, the control input appears in a neighbourhood of the origin, meaning that the Itô stochastic relative degree at zero is defined and equal to two.

We now consider the stochastic system described by (6), where the vector fields f, g, l, m and the map h are given by (10)-(11). We compute the differential of y_t using the Stratonovich rule, yielding

$$dy_t = x_3 dt,$$

hence $y_t^{(1)} = x_3$. The control input does not appear in this expression and therefore the Stratonovich stochastic relative degree, if defined, is higher than one. Differentiating $y_t^{(1)}$ yields

$$dy_t^{(1)} = x_1 dt,$$

or, equivalently, $y_t^{(2)} = x_1$ and the Stratonovich stochastic relative degree, if defined, is higher than two. Finally,

$$dy_t^{(2)} = e^{-x_1}\left[(x_3 - x_2)dt - \cos x_2\, u\, dt + (1 + x_1) \circ d\mathcal{W}_t\right].$$

As the control input appears in the previous expression in a neighbourhood of zero, the Stratonovich relative degree at zero is defined and equal to three.

Example 2. (Larger Itô stochastic relative degree) This second example illustrates the converse case of Example 1. Specifically, the stochastic relative degree of two systems defined by the same differential equation is higher when this is meant in Itô's sense than when it is meant in Stratonovich's sense.

Let $x = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^\top$ and

$$f(x)=\begin{bmatrix} x_1 \\ x_3 - e^{x_1}x_1 - \frac{1}{2}e^{-x_1} \\ x_3 - x_1 \end{bmatrix},\ g(x)=\begin{bmatrix} \cos x_3 \\ -\cos x_3(1+e^{x_1}) \\ -\cos x_3 \end{bmatrix}, \tag{12}$$

$$l(x)=\begin{bmatrix} e^{-x_1} \\ -(1+e^{-x_1}) \\ -e^{-x_1} \end{bmatrix},\ m(x)=\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},\ h(x)=e^{x_1}+x_1+x_2. \tag{13}$$

We consider the system described by (4) and we compute the differential of y_t using Itô's lemma, which yields

$$dy_t = \begin{bmatrix} e^{x_1}+1 & 1 & 0 \end{bmatrix} dx_t + \frac{1}{2} dx_t^\top \begin{bmatrix} e^{x_1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} dx_t.$$

Using Itô's rules on the stochastic differentials, the previous expression boils down to

$$dy_t = (x_1 + x_3)dt,$$

or, equivalently, $y_t^{(1)} = x_1 + x_3$. The stochastic relative degree, if defined, is higher than one and we can differentiate $y_t^{(1)}$ further, yielding

$$dy_t^{(1)} = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix} dx_t = x_3 dt,$$

hence $y_t^{(2)} = x_3$. The stochastic relative degree, if defined, is higher than two and a further differentiation results in

$$dy_t^{(2)} = (x_3 - x_1 - \cos x_3\, u)dt - e^{-x_1} d\mathcal{W}_t.$$

Therefore, in a neighbourhood of the origin the control input appears after differentiating the output three times and we conclude that the Itô stochastic relative degree of the system is defined and equal to three.

We now look at system (6)-(12)-(13). We compute the differential of the output using Stratonovich's differentiation rule, which results in

$$dy_t = \begin{bmatrix} e^{x_1}+1 & 1 & 0 \end{bmatrix} dx_t = (x_1 - \frac{1}{2}e^{-x_1} + x_3)dt,$$

or, equivalently, $y_t^{(1)} = x_1 - \frac{1}{2}e^{-x_1} + x_3$. The Stratonovich stochastic relative degree, if defined, is therefore higher than one and, differentiating further, yields

$$\begin{aligned} dy_t^{(1)} &= \begin{bmatrix} 1 + \frac{1}{2}e^{-x_1} & 0 & 1 \end{bmatrix} dx_t \\ &= x_3 dt + \frac{1}{2} e^{-x_1}(x_1 dt + \cos x_3\, u dt + e^{-x_1} \circ d\mathcal{W}_t). \end{aligned}$$

In a neighbourhood of the origin the control input appears in the second derivative of the output, therefore the Stratonovich stochastic relative degree is defined and equal to two.

Remark 1. Note that the fact that the Itô and Stratonovich stochastic relative degrees of two systems described by the same stochastic differential equations are different is not a paradox. In fact, it is important to notice that (4) and (6) are, indeed, two *different* systems, with different dynamics and, therefore, in general different stochastic relative degrees. However, there exist conversion formulas between the solutions of the two differential equations (4) and (6), see, *e.g.*, [3, Chapter 10]. Thus, by converting one representation into the other using the appropriate conversion formulas, one would obtain the same stochastic relative degree in both conventions.

V. Conclusions

In this note we have recalled the notion of stochastic relative degree in Itô's sense and we have introduced the notion of Stratonovich stochastic relative degree. We have also defined the notion of Stratonovich normal form of a nonlinear stochastic system. We have pointed out the main differences between the two interpretations of relative degree and we have illustrated through examples that, in general, the two stochastic relative degrees are different.

References

[1] B. Øksendal, *Stochastic Differential Equations (Sixth Edition)*. Springer-Verlag, 2003.

[2] J. Yong and X. Y. Zhou, *Stochastic Controls: Hamiltonian Systems and HJB Equations*, ser. Stochastic Modelling and Applied Probability. Springer New York, 1999.

[3] L. Arnold, *Stochastic Differential Equations*, ser. A Wiley-Interscience publication. Wiley, 1974.

[4] E. Wong and M. Zakai, "On the convergence of ordinary integrals to stochastic integrals," *The Annals of Mathematical Statistics*, vol. 36, no. 5, pp. 1560–1564, 10 1965.

[5] A. Isidori, *Nonlinear Control Systems*, ser. Communications and Control Engineering. Springer-Verlag London, 1995.

[6] L. Arnold and P. Imkeller, "Normal forms for stochastic differential equations," *Probability Theory and Related Fields*, vol. 110, no. 4, pp. 559–588, May 1998.

[7] L. Arnold, *Random Dynamical Systems*, ser. Springer Monographs in Mathematics. Springer-Verlag Berlin Heidelberg, 2003.

[8] G. Gaeta and N. Rodríguez Quintero, "Lie-point symmetries and stochastic differential equations," *Journal of Physics A: Mathematical and General*, vol. 32, no. 48, pp. 8485–8505, 1999.

[9] A. Roberts, "Normal form transforms separate slow and fast modes in stochastic dynamical systems," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 1, pp. 12–38, 2008.

[10] A. Mellone and G. Scarciotti, "Normal Form and Exact Feedback Linearisation of Nonlinear Stochastic Systems: the Ideal Case," in *2019 IEEE Conference on Decision and Control (CDC)*, Dec 2019, pp. 3503–3508.

[11] ——, "The Zero Dynamics of Nonlinear Stochastic Systems: Stabilization and Output Tracking in the Ideal Case," in *21st IFAC World Congress (IFAC 2020)*, July 2020, pp. 5053–5058.

[12] ——, "Approximate Feedback Linearisation and Stabilisation of Nonlinear Stochastic Systems," in *21st IFAC World Congress (IFAC 2020)*, July 2020, pp. 5059–5064.

[13] L. Hörmander, *The Analysis of Linear Partial Differential Operators I*, ser. Classics in Mathematics. Springer-Verlag Berlin Heidelberg, 1983.

[14] J. J. Duistermaat and J. A. C. Kolk, *Distributions*, ser. Cornerstones. Birkäuser Basel, 1995.

[15] T. Gard, *Introduction to Stochastic Differential Equations*, ser. Monographs and textbooks in pure and applied mathematics. M. Dekker, 1988.

[16] R. L. Stratonovich, "A new representation for stochastic integrals and equations," *SIAM Journal on Control*, vol. 4, no. 2, pp. 362–371, 1966.